CSCI 6302. Speech Recognition and Synthesis

Time: Fall 2009-2010, Tuesdays and Thursdays from 05:00pm to 06:15pm

Instructor: Daniel Bolaños (daniel.bolanos@colorado.edu)

Description: The course is an introduction to automatic speech recognition and its applications from the Computer Science perspective. Some knowledge of probability and computer programming is preferred.

Contents:

Part I – Speech Recognition

1. Introduction to Speech recognition
    - Speech production and perception
    - Phonetics and phonology

2. Hidden Markov Models (HMMs)

    - Introduction to HMMs
        - Architecture of a HMM
        - Dynamic programming and Dynamic Time Warping
        - Forward and Backward Algorithms
        - Decoding: Viterbi Search
        - Parameter estimation: Baum-Welch Algorithm
    - Density estimation
        - Continuous density
        - Semi-continuous
    - Alternative paradigms for density estimation
        - Artificial Neural Networks
        - Kernel Methods / Support Vector Machines
    - Discriminative training of HMM parameters
        - Maximum Mutual Information (MMI)
        - Minimum Phone Error (MPE)
        - Large Margin Discriminative Training (LMDT)
    - Alternative sequence classification methods:
        - Conditional Random Fields

4. Acoustic Modeling

    - Variability of the speech signal.
    - Measuring speech recognition accuracy
    - Signal processing and feature extraction

- Modeling units
    - Comparison of different units: phonemes, syllables, words...
    - Context dependency
    - Triphone clustering techniques.
- Speaker adaptation
    - Introduction
    - Speaker clustering
    - Maximum A Posteriori (MAP)
    - Maximum Likelihood Linear Regression

5. Language Modeling

    - Probabilistic Context Free Grammar
    - N-gram language models
    - Advanced topics in language modeling

6. Search algorithms for speech recognition

    - From isolated speech recognition to continuous speech recognition
    - Time-Synchronous Viterbi Beam Search
        - Token passing algorithm
        - Decoding network organization: dynamic, static, tree-copy.
        - Pruning strategies: likelihood, histogram and layer dependent pruning.
    - A* decoding.
    - Integrating acoustic and language modeling.
    - Decoding output and evaluation
        - N-best lists
        - Word-graphs
    - Post-processing of recognition hypothesis, multipass strategies.

7. Speech processing topics

    - Confidence estimation and filler models.
    - Dialog systems and VoiceXML
    - Speaker verification
    - Spoken term detection
    - Reference verification
        - Pronunciation error detection
        - Reading assessment

Part II – Speech synthesis

1. Introduction
2. Synthesis techniques based on vocal tract models
3. Synthesis by concatenation and signal processing modification

4. HMM based synthesis
5. Synthesis based on unit selection

Required work:

- Readings: will be provided to students to be read in preparation for the class.

- Project: at the beginning of the course a list of project suggestions will be presented to the students. Projects will be connected to different topics of speech processing (mostly practical) and will be accompanied with appropriate references to the literature. Students will choose the project they want to do before the end of the second week of the course. Projects will be carried out individually and exceptionally by teams of two students. To make it easier to keep up with the assignment, there will be deliverables every two weeks. Deliverables will be updates in the form of document progress. The project will be presented orally the last week of the course. Additionally the student will summarize the work done in a four page document following the format of a research article.

Textbooks:

- Spoken Language Processing: A Guide to Theory, Algorithm and System Development, by Xuedong Huang, Alex Acero and Hsiao-Wuen Hon.
- Fundamentals of Speech Recognition, by Lawrence Rabiner and Biing-Hwang Juang. Prentice Hall, Signal Processing Series.
- Text to Speech Synthesis, by Paul Taylor. Cambridge University Press, 1 edition. March 2, 2009.
- Speech and Language Processing, by Jurafsky and Martin. Prentice-Hall, 2nd edition, 2008.
- Research articles that will complement the lessons.